## Universidad del Rosario

**Escuela de ingeniería ciencia y tecnología Proyecto Ingeniería de datos**

## Rosa Galicia, Juan David Urbano, Karol Torres. Introducción:

El conjunto de datos con los que se trabajó está basado en los medicamentos vitales que no están

disponibles o están escasos en nuestro país.

Ya que hay medicamentos que son indispensables en algunos casos clínicos para salvaguardar la vida de los pacientes o mejorar su estado de salud, pero en algunos casos por diversos factores no se están disponibles o las cantidades no son suficientes.

La entidad que suministró la información fue el Instituto Nacional de Vigilancia de Medicamentos y Alimentos y su última actualización fue el 9 de octubre de 2023

Cabe resaltar que el conjunto de datos seleccionado cuenta con 5.930 filas y 16 columnas, de las cuales obviamos algunas por ambigüedad, insignificancia o incluso redundancia en los datos.

El link es el siguiente: [https://www.datos.gov.co/Salud-y-Protecci-n-Social/MEDICAMENTOS-](https://www.datos.gov.co/Salud-y-Protecci-n-Social/MEDICAMENTOS-VITALES-NO-DISPONIBLES/sdmr-tfmf) [VITALES-NO-DISPONIBLES/sdmr-tfmf](https://www.datos.gov.co/Salud-y-Protecci-n-Social/MEDICAMENTOS-VITALES-NO-DISPONIBLES/sdmr-tfmf).

A continuación, se presenta una descripción general de las columnas utilizadas en el presente proyecto:

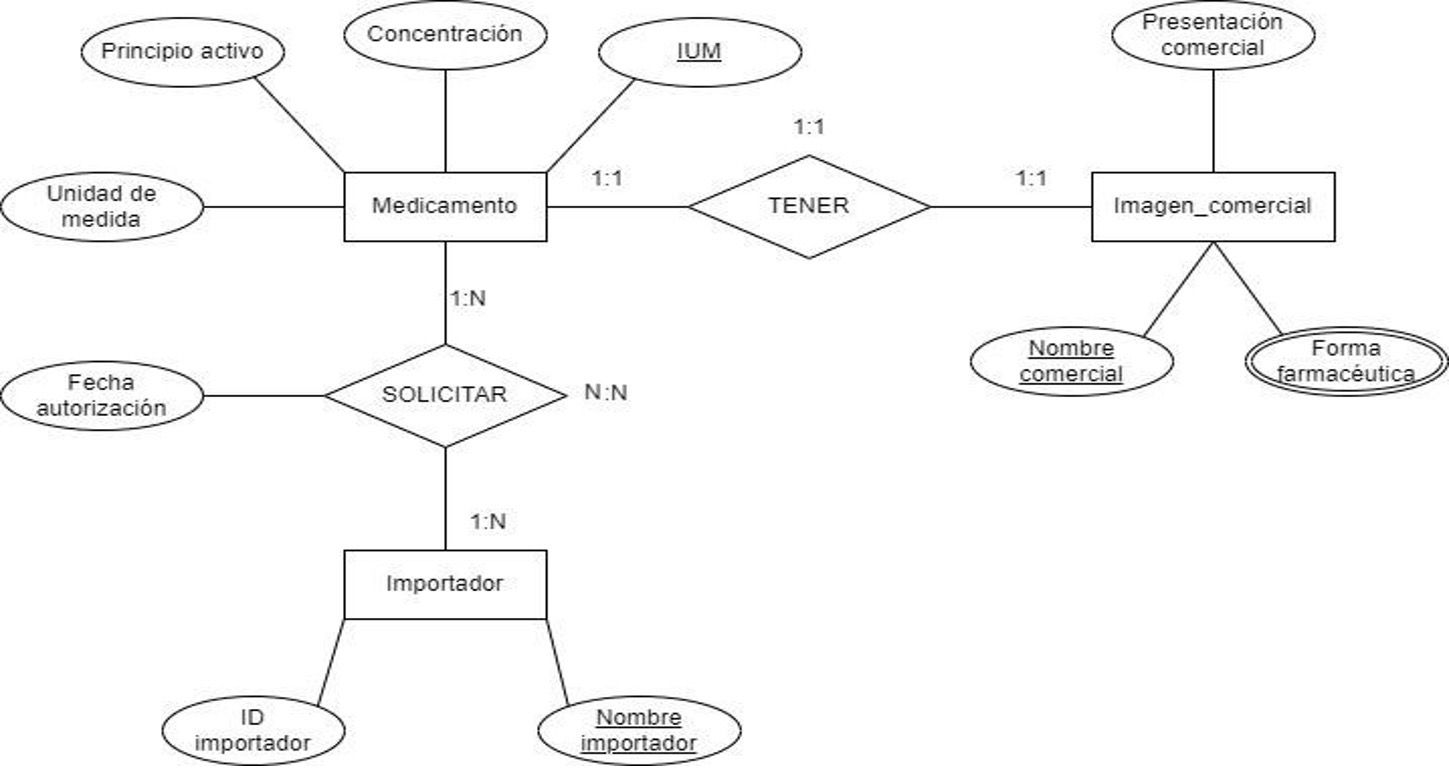
|  |  |
| --- | --- |
| **COLUMNAS** | **DESCRIPCIÓN** |
| **FECHA\_DE\_AUTORIZACIÓN** | Fecha en que se generó la autorización de la solicitud del medicamento |
| **TIPO\_DE\_SOLICITUD** | Se registra si el trámite autorizado fue para un paciente especifico o para varios pacientes. |
| **SOLICITANTE/IMPORTADOR** | Nombre de la empresa importador a quien se le emite la autorización de importación |
| **IUM** | Identificador único del medicamento, acorde a la Resolución 3166 de 2015 |
| **PRINCIPIO\_ACTIVO** | Es el componente o ingrediente principal de un medicamento. |
| **CONCENTRACIÓN\_DEL\_MEDICAMENTO** | Relación entre la cantidad de soluto y la cantidad de disolución o de disolvente. |
| **UNIDAD DE MEDIDA** | Cantidad estandarizada expresada en el sistema internacional de medida |
| **FORMA\_FARMACÉUTICA** | Forma en la cual se presenta un producto  farmacéutico terminado que permite su dosificación y aplicación terapéutica |
| **NOMBRE\_COMERCIAL** | Nombre con el que comercializa el producto |
| **PRESENTACIÓN\_COMERCIAL** | Número de unidades de contenido en un mismo empaque |
| **CANTIDAD\_SOLICITADA** | Número de unidades aprobadas |

|  |  |
| --- | --- |
| **PRESENTACIÓN\_COMERCIAL** | Número de unidades de contenido en un mismo empaque |
| **DIAGNOSTICO** | Enfermedad o patología para la que indicado el medicamento autorizado |
| **CIE-10** | Código según la clasificación internacional de enfermedades |

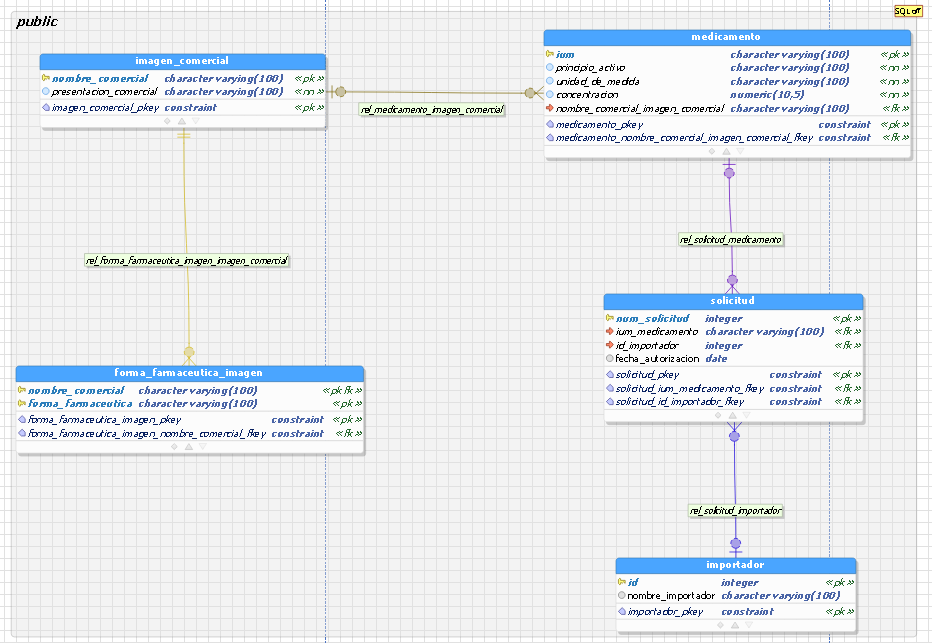
# Reglas de negocio

* Las entidades encargadas de importar los medicamentos llevan control de aquellos fármacos que se hayan solicitado y aun no estén disponibles.
* Cada entidad registra los siguientes datos tras la autorización de la solicitud en espera: fecha/hora de autorización, tipo de solicitud, cantidad solicitada, unidad de medida, diagnóstico y CIE10 del diagnóstico y nombre del importador.
* La concentración del medicamento puede ser expresada en peso o volumen.
* Cada medicamento tiene un principio activo, una concentración y su IUM correspondiente
* El IUM es el Identificador Único de cada Medicamento.
* Adicionalmente cada medicamento tiene una imagen comercial con la que esta estandarizado en el mercado con su respectivo nombre comercial, forma farmacéutica y presentación comercial.
* El importador registra la fecha/hora en que se le dio la autorización a la solicitud de cada medicamento
* Puede que la unidad de medida de un medicamento varie según la empresa que lo registró
* Todos los medicamentos tienen un nombre único comercialmente.
* Pueden existir varios medicamentos de marca con el mismo principio activo.
* Cada solicitud tiene un número que la identifica y una fecha en que se autorizó el pedido de medicamento.

# Modelo Entidad relación:



**Modelo relacional (Pg\_modeler):**



## Verificación normalización Primera Normal (1FN)

* Se observa que ninguna tabla tiene una columna que genere redundancia en los datos y cada columna tiene un valor INDIVISIBLE

## Segunda Normal (2FN)

* + Existe una dependencia funcional completa en todas las tablas

## Tercera forma normal (3FN)

* + Todos los atributos que no son llave primaria dependen únicamente de la llave primaria y no de otro atributo (no hay dependencias transitivas)

Después de hacer el proceso de normalización hasta la (3FN) se concluye que no hay modificación en el modelo relacional.

## DASHBOARDS

## Creamos cuatro escenarios de análisis de las gráficas

## 1.Rango de costos en el que se encuentra cada medicamento

## 2.Importadores que más solicitudes realizaron

## 3.Medicamentos con mayor demanda

## 4. Frecuencia de cada tipo de solicitud mediante número de solicitudes

## Rango de costos de cada medicamento

## Gráfico, Gráfico de líneas Descripción generada automáticamente

## Gráfico, Gráfico circular Descripción generada automáticamente

## En estas gráficas podemos evidenciar el rango de costos en el que se encuentran los medicamentos.

## En el gráfico 1, el eje x representa los intervalos de costos, mientras que el eje Y representa la cantidad de medicamentos. Se puede observar un patrón inversamente proporcional, algo que no se refleja de manera evidente en el gráfico circular. Podemos apreciar que los medicamentos más solicitados se sitúan en los rangos de costos de [10-20), [20-30) y [30-40). Por el contrario, aquellos con una demanda más baja se encuentran en los intervalos de [40-50) y [50-60).

## En el segundo gráfico se muestran los porcentajes del intervalo de costos de los medicamentos, esta gráfica es básicamente un complemento de la primera gráfica

## DISCUSIÓN: Una de las ventajas de nuestro polígono de frecuencias es que proporciona una representación visual clara en donde se facilita la identificación de las relaciones que necesitamos; por otro lado, puede ser sensible a valores atípicos ya que se puede distorsionar la percepción de la distribución general. Nuestro diagrama circular se ve claro, pero al mismo tiempo no podemos identificar tan fácilmente la relación cantidad vs costo.

## Importadores con más solicitudes

## 

## 

## En este caso podemos observar los importadores que hicieron más de 80 solicitudes.

## Se decidió tomar los mayores a 80 debido a que había varios importadores que solo pedían una solicitud y generaba que la escala del eje no fuera muy exacta.

## Aunque los dos gráficos podrían ser redundantes, el primer gráfico proporciona una representación más visual de la cantidad aproximada, mientras que el segundo grafico expone la información de manera más precisa mediante la presentación porcentual.

## Nos podemos percatar de que la empresa que presenta más solicitudes es Audifarma S.A, puede ser debido a que esta empresa dispensa medicamentos a varias de las principales E.P.S. del país, como lo es Nueva E.P.S., SaludTotal, Coosalud, Compensar y la recién añadida E.P.S Sanitas, por lo cual no se puede determinar si existe o no un monopolio, pero en el caso de que si existiese, nos damos cuenta de que las E.P.S´s perjudicadas son las más grandes e importantes del país.

## DISCUSIÓN: Nuestro diagrama de barras suele verse claro y con la información organizada, pero en algunos casos si tenemos más cantidad de importadores con esta gráfica se van a superponer y va a ser imposible su identificación, de igual manera con nuestro diagrama circular al tener varios nombres de importadores nos podemos confundir con los porcentajes de cada variable.

## Medicamentos con mayor demanda

## 

## Respecto a estas graficas decidimos tomar los medicamentos que tenían más de 50 solicitudes para tener una gráfica más precisa. Con lo cual, podemos ver que el medicamento “Translarna” es el medicamento que tiene mayor solicitud y tiene menos disponibilidad. Ahora bien, nos damos cuenta de que está en el rango de precios más altos de los medicamentos en donde posiblemente tiene alguna relación con los costos de su producción.

## DISCUSIÓN: Nuestra gráfica de barras se ve clara y podemos reconocer la cantidad de cada medicamento desde el más solicitado al menos solicitado; en nuestra gráfica circular también podemos evidenciar muy fácilmente el porcentaje de cada medicamento con su respectivo nombre comercial.

## Frecuencia de cada tipo de solicitud mediante número de solicitudes

## Gráfico Descripción generada automáticamente

## Gráfico, Gráfico circular Descripción generada automáticamenteFinalmente, podemos ver por medio de estas gráficas el tipo de solicitud de todos los medicamentos lo cual sería muy útil al momento de clasificar la importación de estos, también identificamos que los medicamentos mayormente solicitados son para más de un paciente, luego para urgencia clínica y los medicamentos con menor solicitud son para un paciente especifico, esto se debe a que se trata de pacientes con afecciones poco comunes.

## DISCUSIÓN: Nuestra grafica de barras en este caso al ser pocas variables es más fácil reconocer cada tipo de solicitud, una de sus desventajas es que al tener unas grandes cantidades de medicamentos no podemos ver exactamente su valor. En la gráfica circular se reconoce muy fácilmente el porcentaje de cada tipo de solicitud al ser pocas variables.

## INTRUCCIONES PARA VISUALIZAR GRÁFICAS

## En nuestro caso para correr de manera eficiente el Dashboard.py fue necesario ejecutarlo desde el símbolo de sistema CMD, y ya nos retornaba el link correspondiente a la página HTML, para evitar inconvenientes con la ejecución desde un entorno como lo es Spyder, es recomendable ejecutarlo desde el CMD, indicando la ruta de acceso y el comando “python Dashboard.py”.

## CONCLUSIONES:

Durante el desarrollo de este proyecto nos vimos enfrentados a varias dificultades que contribuyeron de manera negativa en las entregas del mismo. La primera entrega nos enfrentó a dificultades iniciales en la búsqueda de una base de datos óptima, lo que afectó la calidad de nuestra presentación inicial. Aprendimos la importancia de una selección cuidadosa de datos y la necesidad de adaptarnos a situaciones imprevistas. La corrección y ajuste en la segunda entrega nos enseñaron sobre la importancia de la coordinación efectiva y la gestión de archivos en GitHub. La transformación de datos y la iteración en el diseño del diagrama de entidad-relación nos desafiaron a repensar nuestra aproximación al proyecto. A pesar de las dificultades, logramos construir dashboards informativos que representan el rango de costos de medicamentos, los importadores más activos, los medicamentos con mayor demanda y la frecuencia de cada tipo de solicitud con el fin de satisfacer nuestros escenarios de análisis iniciales. Al construir el diagrama relacional normalizado en tercera forma normal, nos enfrentamos a complicaciones específicas en la implementación del código SQL. La normalización, aunque esencial, resultó ser un proceso más complejo de lo previsto, especialmente al lidiar con las dependencias funcionales y asegurar la integridad referencial. Aprendimos a abordar eficientemente estos desafíos mediante una cuidadosa planificación y revisión del código SQL.

La carga de datos en PostgreSQL también presentó desafíos, ya que debíamos asegurarnos de que todos los datos se insertaran correctamente y cumplieran con las restricciones de la base de datos. La identificación y descripción de posibles escenarios de análisis nos obligaron a escribir consultas SQL complejas para extraer la información relevante de manera eficiente.

Al desarrollar los módulos en Python para la conexión a la base de datos y la creación de un repositorio GIT presentaron desafíos adicionales. A pesar de estos desafíos, el proceso nos proporcionó valiosas lecciones sobre la importancia de la planificación cuidadosa, la revisión de código y la resolución creativa de problemas en entornos SQL y Python. Estas habilidades son fundamentales para enfrentar futuros proyectos con confianza y eficacia.

## URL de Github

https://github.com/juanda14254/Base\_datos\_Medicamentos